Joint Sparse Representation for Robust Multimodal Biometrics Recognition

Sumit Shekhar, *Student Member, IEEE,* Vishal M. Patel, *Member, IEEE,* Nasser M. Nasrabadi, *Fellow, IEEE,* and Rama Chellappa, *Fellow, IEEE*

.

***Abstract*—Traditional biometric recognition systems rely on a single biometric signature for authentication. While the advantage of using multiple sources of information for establishing the identity has been widely recognized, computational models for multimodal biometrics recognition have only recently received attention. We propose a novel multimodal multivariate sparse representation method for multimodal biometrics recognition, which represents the test data by a sparse linear combination of training data, while constraining the observations from different modalities of the test subject to share their sparse representations. Thus, we simultaneously take into account correlations as well as coupling information between biometric modalities. We modify our model so that it is robust to noise and occlusion. A multimodal quality measure is also proposed to weigh each modality as it gets fused. Furthermore, we also kernelize the algorithm to handle non-linearity in data. The optimization problem is solved using an efficient alternative direction method. Various experiments show that our method compares favorably with competing fusion-based methods.**

***Index Terms*—Multimodal biometrics, feature fusion, sparse representation.**

I. Introduction

Unimodal biometric systems rely on a single source of information such as a single iris or fingerprint or face for authentication [1]. Unfortunately these systems have to deal with some of the following inevitable problems [2]: (a) Noisy data: poor lighting of a user's face or an iris image with contact lense are examples of noisy data. (b) Non-universality: the biometric system based on a single source of evidence may not be able to capture meaningful data from some users. For instance, an iris biometric system may extract incorrect texture patterns from the iris of certain users due to the presence of contact lenses. (c) Intra-class variations: in the case of fingerprint recognition wrinkles due to wet fingers [3] can cause these variations. These types of variations often occur when a user incorrectly interacts with the sensor. (d) Spoof attack: hand signature forgery is an example of this type of attack. It has been observed that some of the limitations of unimodal biometric systems can be addressed by deploying multimodal biometric systems that essentially integrate the evidence presented by multiple sources of information such as iris, fingerprints and face. Such systems are less vulnerable to spoof attacks as it would be difficult for an imposter to simultaneously spoof multiple biometric traits of a genuine user. Due to sufficient population coverage, these systems are able to address the problem of non-universality.

Sumit Shekhar, Vishal M. Patel and R. Chellappa are with the Department of Electrical and Computer Engineering and the Center for Automation Research, UMIACS, University of Maryland, College Park, MD 20742 USA (e-mail: { sshekha,pvishalm,rama }@umiacs.umd.edu)

Nasser M. Nasrabadi is with the U.S. Army Research Lab, Adelphi, MD 20783 USA (e-mail: nasser.m.nasrabadi@us.army.mil).

Classification in multibiometric systems is done by fusing information from different biometric modalities. The information fusion can be done at different levels, which can be broadly divided into feature level, score level and rank/decision level fusion. Due to preservation of raw information, feature level fusion can be more discriminative than score or decision level fusion [4]. But, there have been very little effort in exploring feature level fusion in the biometric community. This is because of the different output formats of different sensors, which result in features with different dimensions. Often the features have large dimensions, and fusion becomes difficult at feature level. The prevelant method is feature concatenation, which has been used for different multibiometric settings [5]–[7]. However, in many scenarios, each modality produces high-dimensional features. In such cases, the method is both impractical and non-robust. It also cannot exploit the constraint that features of different modalities should share the same identity.

In recent years, theories of Sparse Representation (SR) and Compressed Sensing (CS) have emerged as powerful tools for efficient processing of data in non-traditional ways [8]. This has led to a resurgence in interest in the principles of SR and CS for biometrics recognition [9]. Wright *et al.* [10] proposed a robust sparse representation-based classification (SRC) algorithm for face recognition. It was shown that by exploiting the inherent sparsity of data, one can obtain improved recognition performance over traditional methods especially when the data is contaminated by various artifacts such as illumination variations, disguise, occlusion and random pixel corruption. Pillai *et al.* extended this work for robust cancelable iris recognition in [11]. Nagesh and Li [12] presented an expression-invariant face recognition method using distributed compressed sensing and joint sparsity models. Patel *et al.* [13] proposed dictionary based method for face recognition under varying pose and illumination. A discriminative dictionary learning method for face recognition was also proposed by Zhang and Li [14]. For a survey of applications of SR and CS algorithms to biometric recognition, see [8], [9], [15], [16] and the references therein.

Motivated by the success of SR in unimodal biometric recognition, we propose a joint sparsity-based algorithm for

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
DEC 2012		**00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Joint Sparse Representation for Robust Multimodal Biometrics Recognition	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Maryland, College Park,Department of Electrical and Computer Engineering and the Center for Automation ,the Center for Automation Research ,College Park,MD,20742	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Traditional biometric recognition systems rely on a single biometric signature for authentication. While the advantage of using multiple sources of information for establishing the identity has been widely recognized, computational models for multimodal biometrics recognition have only recently received attention. We propose a novel multimodal multivariate sparse representation method for multimodal biometrics recognition which represents the test data by a sparse linear combination of training data, while constraining the observations from different modalities of the test subject to share their sparse representations. Thus, we simultaneously take into account correlations as well as coupling information between biometric modalities. We modify our model so that it is robust to noise and occlusion. A multimodal quality measure is also proposed to weigh each modality as it gets fused. Furthermore, we also kernelize the algorithm to handle non-linearity in data. The optimization problem is solved using an efficient alternative direction method. Various experiments show that our method compares favorably with competing fusion-based methods.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**13**	

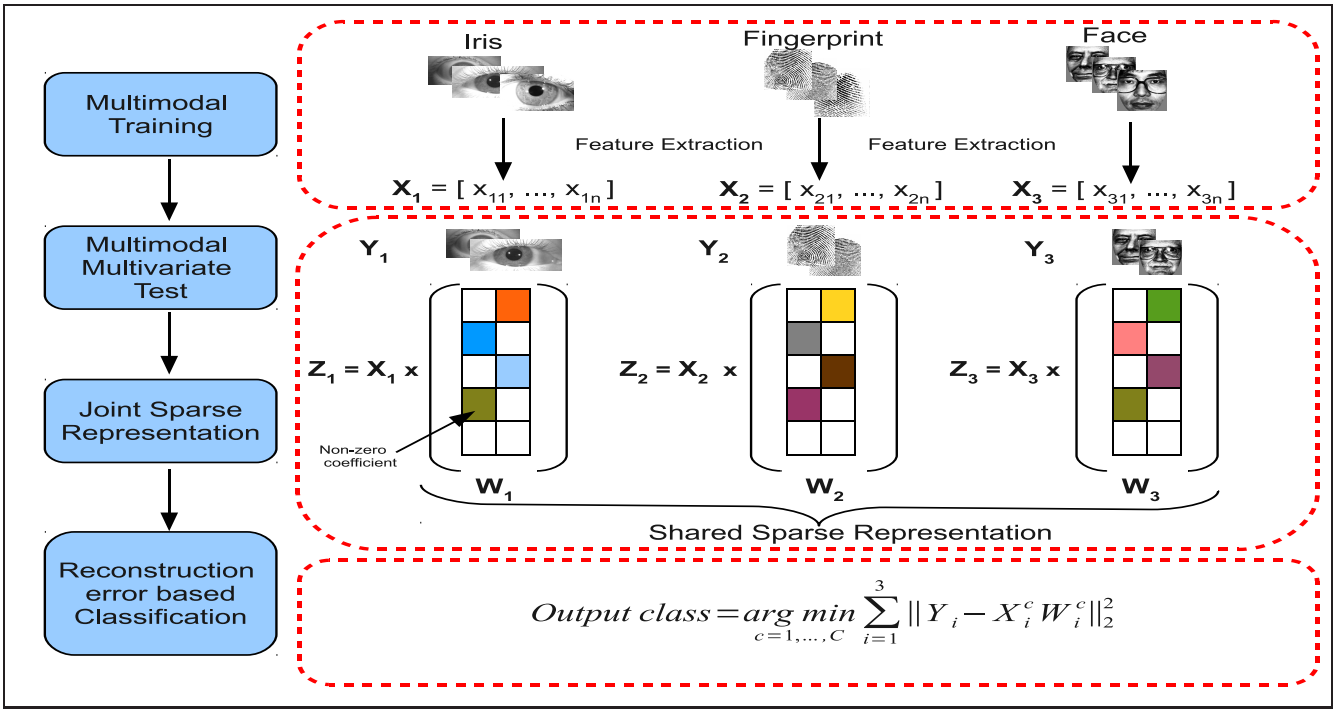

Fig. 1: Overview of our algorithm.

multimodal biometrics recognition [1]. Figure 1 presents an overview of our framework. It is based on the well known regularized regression method, multi-task multivariate Lasso [18], [19]. Our method imposes common sparsities both within each biometric modality and across different modalities. Furthermore, we extend our model so that it can deal with both occlusion and noise. Note that our method is different from some of the previously proposed classification algorithms based on joint sparse representation. Yuan and Yan [20] proposed a multi task sparse linear regression model for image classification. This method uses group sparsity to combine different features of an object for classification. Zhang *et al.* [21] proposed a joint dynamic sparse representation model for object recognition. Their essential goal was to recognize the same object viewed from multiple observations i.e., different poses. Our method is more general in that it does not only considers multivariate sparse representations but it can also deal with multitask multivariate sparse representations which are natural in multimodal biometrics. One of the key features of our model is that it can deal with both occlusion and noise. Furthermore, using kernel methods, we present non-linear extensions of our joint sparse representation method.

This paper makes the following contributions:

- We present a robust feature level fusion algorithm for multibiometric recognition. Through the proposed joint sparse framework, we can easily handle different dimensions of different modalities by forcing different features to interact through their sparse coefficients. Furthermore, the proposed algorithm can efficiently handle large dimensional feature vectors.
- We make the classification robust to occlusion and noise by introducing an error term into the optimization framework.
- The algorithm is easily generalizable to handle multiple test inputs from a modality.
- We introduce a quality measure for multimodal fusion based on joint sparse representation.
- Lastly, we kernelize the algorithm to handle non-linearity in data samples.

[1]A preliminary version of this work appeared in [17].

A. Paper Organization

The paper is organized in five sections. In section II, we describe the proposed sparsity-based multimodal recognition algorithm which is extended to kernel case in section IV. The quality measure is described in III. Experimental evaluations on a comprehensive multimodal dataset and a face database have been described in section V. Finally, in section VI, we describe the computational complexity. Concluding remarks are presented in section VII.

II. Joint Sparsity-based Multimodal Biometrics Recognition

Consider a multimodal C-class classification problem with D different biometric traits. Suppose there are p_i training samples in each biometric trait. For each biometric trait $i = 1, \ldots, D$, we denote

$$\mathbf{X}^i = [\mathbf{X}^i_1, \mathbf{X}^i_2, \ldots, \mathbf{X}^i_C]$$

as an $n_i \times p_i$ dictionary of training samples consisting of C sub-dictionaries $\mathbf{X}^i_k$'s corresponding to C different classes. Each sub-dictionary

$$\mathbf{X}^i_j = [\mathbf{x}^i_{j,1}, \mathbf{x}^i_{j,2}, \ldots, \mathbf{x}^i_{j,p_j}] \in \mathbb{R}^{n \times p_j}$$

represents a set of training data from the ith modality labeled with the jth class. Note that n_i is the feature dimension of each sample and there are p_j number of training samples in class j. Hence, there are a total of $p = \sum_{j=1}^{C} p_j$ many samples in the dictionary $\mathbf{X}_C^i$. Elements of the dictionary are often referred to as atoms. In multimodal biometrics recognition problem, given a test samples (matrix) $\mathbf{Y}$, which consists of D different modalities $\{\mathbf{Y}^1, \mathbf{Y}^2, \ldots, \mathbf{Y}^D\}$ where each sample $\mathbf{Y}^i$ consists of d_i observations $\mathbf{Y}^i = [\mathbf{y}_1^i, \mathbf{y}_2^i, \ldots, \mathbf{y}_d^i] \in \mathbb{R}^{n \times d_i}$, the objective is to identify the class to which a test sample $\mathbf{Y}$ belongs to. In what follows, we present a multimodal multivariate sparse representation-based algorithm for this problem [18], [19], [22].

A. Multimodal multivariate sparse representation

We want to exploit the joint sparsity of coefficients from different biometrics modalities to make a joint decision. To simplify this model, let us consider a bimodal classification problem where the test sample $\mathbf{Y} = [\mathbf{Y}^1, \mathbf{Y}^2]$ consists of two different modalities such as iris and face. Suppose that $\mathbf{Y}^1$ belongs to the jth class. Then, it can be reconstructed by a linear combination of the atoms in the sub-dictionary $\mathbf{X}_j^1$. That is, $\mathbf{Y}^1 = \mathbf{X}^1\mathbf{\Gamma}^1 + \mathbf{N}^1$, where $\mathbf{\Gamma}^1$ is a sparse matrix with only p_j nonzero rows associated with the jth class and $\mathbf{N}^1$ is the noise matrix. Similarly, since $\mathbf{Y}^2$ represents the same subject, it belongs to the same class and can be represented by training samples in $\mathbf{X}_j^2$ with different set of coefficients $\mathbf{\Gamma}_j^2$. Thus, we can write $\mathbf{Y}^2 = \mathbf{X}^2\mathbf{\Gamma}^2 + \mathbf{N}^2$, where $\mathbf{\Gamma}^2$ is a sparse matrix that has the same sparsity pattern as $\mathbf{\Gamma}^1$. If we let $\mathbf{\Gamma} = [\mathbf{\Gamma}^1, \mathbf{\Gamma}^2]$, then $\mathbf{\Gamma}$ is a sparse matrix with only p_j nonzeros rows.

In the more general case where we have D modalities, if we denote $\{\mathbf{Y}^i\}_{i=1}^D$ as a set of D observations each consisting of d_i samples from each modality and let $\mathbf{\Gamma} = [\mathbf{\Gamma}^1, \mathbf{\Gamma}^2, \ldots, \mathbf{\Gamma}^D] \in \mathbb{R}^{p \times d}$ be the matrix formed by concatenating the coefficient matrices with $d = \sum_{i=1}^{D} d_i$, then we can seek for the row-sparse matrix $\mathbf{\Gamma}$ by solving the following ℓ_1/ℓ_q-regularized least square problem

$$\hat{\mathbf{\Gamma}} = \arg\min_{\mathbf{\Gamma}} \frac{1}{2} \sum_{i=1}^{D} \|\mathbf{Y}^i - \mathbf{X}^i\mathbf{\Gamma}^i\|_F^2 + \lambda \|\mathbf{\Gamma}\|_{1,q} \tag{1}$$

where λ is a positive parameter and q is set greater than 1 to make the optimization problem convex. Here, $\|\mathbf{\Gamma}\|_{1,q}$ is a norm defined as $\|\mathbf{\Gamma}\|_{1,q} = \sum_{k=1}^{p} \|\boldsymbol{\gamma}^k\|_q$ where $\boldsymbol{\gamma}^k$'s are the row vectors of $\mathbf{\Gamma}$ and $\|\mathbf{Y}\|_F$ is the Frobenius norm of the matrix $\mathbf{Y}$ defined as $\|\mathbf{Y}\|_F = \sqrt{\sum_{i,j} Y_{i,j}^2}$. Once $\hat{\mathbf{\Gamma}}$ is obtained, the class label associated with an observed vector is then declared as the one that produces the smallest approximation error.

$$\hat{j} = \arg\min_{j} \sum_{i=1}^{D} \|\mathbf{Y}^i - \mathbf{X}^i\boldsymbol{\delta}_j^i(\mathbf{\Gamma}^i)\|_F^2, \tag{2}$$

where $\boldsymbol{\delta}_j^i$ is the matrix indicator function defined by keeping rows corresponding to the jth class and setting all other rows equal to zero. Note that the optimization problem (1) reduces to the conventional Lasso [23] when $D = 1$ and $d = 1$. In the case, when $D = 1$ (1) is referred to as multivariate Lasso [18].

B. Robust multimodal multivariate sparse representation

In this section, we consider a more general problem where the data is contaminated by noise. In this case, the observation model can be modeled as

$$\mathbf{Y}^i = \mathbf{X}^i\mathbf{\Gamma}^i + \mathbf{Z}^i + \mathbf{N}^i, \quad i = 1, \ldots D, \tag{3}$$

where $\mathbf{N}^i$ is a small dense additive noise and $\mathbf{Z}^i \in \mathbb{R}^{n \times d_i}$ is a matrix of background noise (occlusion) with arbitrarily large magnitude. One can assume that each $\mathbf{Z}^i$ is sparsely represented in some basis $\mathbf{B}^i \in \mathbb{R}^{n \times m^i}$. That is, $\mathbf{Z}^i = \mathbf{B}^i\mathbf{\Lambda}^i$ for some sparse matrices $\mathbf{\Lambda}^i \in \mathbb{R}^{m_i \times d_i}$. Hence, (3) can be rewritten as

$$\mathbf{Y}^i = \mathbf{X}^i\mathbf{\Gamma}^i + \mathbf{B}^i\mathbf{\Lambda}^i + \mathbf{N}^i, \quad i = 1, \ldots D, \tag{4}$$

With this model, one can simultaneously recover the coefficients $\mathbf{\Gamma}^i$ and $\mathbf{\Lambda}^i$ by taking advantage of the fact that $\mathbf{\Lambda}^i$ are sparse

$$\hat{\mathbf{\Gamma}}, \hat{\mathbf{\Lambda}} = \arg\min_{\mathbf{\Gamma}, \mathbf{\Lambda}} \frac{1}{2} \sum_{i=1}^{D} \|\mathbf{Y}^i - \mathbf{X}^i\mathbf{\Gamma}^i - \mathbf{B}^i\mathbf{\Lambda}^i\|_F^2 + \lambda_1 \|\mathbf{\Gamma}\|_{1,q} + \lambda_2 \|\mathbf{\Lambda}\|_1, \tag{5}$$

where λ_1 and λ_2 are positive parameters and $\mathbf{\Lambda} = [\mathbf{\Lambda}^1, \mathbf{\Lambda}^2, \ldots, \mathbf{\Lambda}^D]$ is the sparse coefficient matrix corresponding to occlusion. The ℓ_1-norm of matrix $\mathbf{\Lambda}$ is defined as $\|\mathbf{\Lambda}\|_1 = \sum_{i,j} |\Lambda_{i,j}|$. Note that the idea of exploiting the sparsity of occlusion term has been studied by Wright *et al.* [10] and Candes *et al.* [24].

Once $\mathbf{\Gamma}, \mathbf{\Lambda}$ are computed, the effect of occlusion can be removed by setting $\tilde{\mathbf{Y}}^i = \mathbf{Y}^i - \mathbf{B}^i\mathbf{\Lambda}^i$. One can then declare the class label associated to an observed vector as

$$\hat{j} = \arg\min_{j} \sum_{i=1}^{D} \|\mathbf{Y}^i - \mathbf{X}^i\boldsymbol{\delta}_j^i(\mathbf{\Gamma}^i) - \mathbf{B}^i\mathbf{\Lambda}^i\|_F^2. \tag{6}$$

C. Optimization algorithm

In this section, we present an algorithm to solve (5) based on the classical alternating direction method of multipliers (ADMM) [25], [26]. Note that the optimization problem (1) can be solved by setting λ_2 equal to zero. Let

$$\mathcal{C}(\mathbf{\Gamma}, \mathbf{\Lambda}) = \frac{1}{2} \sum_{i=1}^{D} \|\mathbf{Y}^i - \mathbf{X}^i\mathbf{\Gamma}^i - \mathbf{B}^i\mathbf{\Lambda}^i\|_F^2.$$

Then, our goal is to solve the following optimization problem

$$\min_{\mathbf{\Gamma}, \mathbf{\Lambda}} \mathcal{C}(\mathbf{\Gamma}, \mathbf{\Lambda}) + \lambda_1 \|\mathbf{\Gamma}\|_{1,q} + \lambda_2 \|\mathbf{\Lambda}\|_1. \tag{7}$$

In ADMM the idea is to decouple $\mathcal{C}(\mathbf{\Gamma}, \mathbf{\Lambda})$, $\|\mathbf{\Gamma}\|_{1,q}$ and $\|\mathbf{\Lambda}\|_1$ by introducing auxiliary variables to reformulate the problem into a constrained optimization problem

$$\min_{\mathbf{\Gamma}, \mathbf{\Lambda}, \mathbf{U}, \mathbf{V}} \mathcal{C}(\mathbf{\Gamma}, \mathbf{\Lambda}) + \lambda_1 \|\mathbf{V}\|_{1,q} + \lambda_2 \|\mathbf{U}\|_1 \quad \text{s. t.} \quad \mathbf{\Gamma} = \mathbf{V}, \mathbf{\Lambda} = \mathbf{U}. \tag{8}$$

Since, (8) is an equally constrained problem, the Augmented Lagrangian method (ALM) [25] can be used to solve the

problem. This can be done by minimizing the augmented lagrangian function $f_{\alpha_\Gamma,\alpha_\Lambda}(\mathbf{\Gamma},\mathbf{\Lambda},\mathbf{V},\mathbf{U};\mathbf{A}_\Lambda,\mathbf{A}_\Gamma)$ defined as

$$\mathcal{C}(\mathbf{\Gamma},\mathbf{\Lambda})+\lambda_2\|\mathbf{U}\|_1+\langle\mathbf{A}_\Lambda,\mathbf{\Lambda}-\mathbf{U}\rangle+\frac{\alpha_\Lambda}{2}\|\mathbf{\Lambda}-\mathbf{U}\|_F^2+\lambda_1\|\mathbf{V}\|_{1,q}+\langle\mathbf{A}_\Gamma,\mathbf{\Gamma}-\mathbf{V}\rangle+\frac{\alpha_\Gamma}{2}\|\mathbf{\Gamma}-\mathbf{V}\|_F^2, \quad (9)$$

where $\mathbf{A}_\Lambda$ and $\mathbf{A}_\Gamma$ are the multipliers of the two linear constraints, and $\alpha_\Lambda,\alpha_\Gamma$ are the positive penalty parameters. The ALM algorithm solves $f_{\alpha_\Gamma,\alpha_\Lambda}(\mathbf{\Gamma},\mathbf{\Lambda},\mathbf{V},\mathbf{U};\mathbf{A}_\Lambda,\mathbf{A}_\Gamma)$ with respect to $\mathbf{\Gamma},\mathbf{\Lambda},\mathbf{U}$ and $\mathbf{V}$ jointly, keeping $\mathbf{A}_\Gamma$ and $\mathbf{A}_\Lambda$ fixed and then updating $\mathbf{A}_\Gamma$ and $\mathbf{A}_\Lambda$ keeping the remaining variables fixed. Due to the separable structure of the objective function $f_{\alpha_\Gamma,\alpha_\Lambda}$, one can further simplify the problem by minimizing $f_{\alpha_\Gamma,\alpha_\Lambda}$ with respect to variables $\mathbf{\Gamma},\mathbf{\Lambda},\mathbf{U}$ and $\mathbf{V}$, separately. Different steps of the algorithm are given in Algorithm 1. In what follows, we describe each of the suboptimization problems in detail.

Algorithm 1: Alternating Direction Method of Multipliers (ADMM).

Initialize: $\mathbf{\Gamma}_0,\mathbf{U}_0,\mathbf{V}_0,\mathbf{A}_{\Lambda,0},\mathbf{A}_{\Gamma,0},\alpha_\Gamma,\alpha_\Lambda$
While not converged do
1. $\mathbf{\Gamma}_{t+1}=\arg\min_{\mathbf{\Gamma}} f_{\alpha_\Gamma,\alpha_\Lambda}(\mathbf{\Gamma},\mathbf{\Lambda}_t,\mathbf{U}_t,\mathbf{V}_t;\mathbf{A}_{\Gamma,t},\mathbf{A}_{\Lambda,t})$
2. $\mathbf{\Lambda}_{t+1}=\arg\min_{\mathbf{\Lambda}} f_{\alpha_\Gamma,\alpha_\Lambda}(\mathbf{\Gamma}_{t+1},\mathbf{\Lambda},\mathbf{U}_t,\mathbf{V}_t;\mathbf{A}_{\Gamma,t},\mathbf{A}_{\Lambda,t})$
3. $\mathbf{U}_{t+1}=\arg\min_{\mathbf{U}} f_{\alpha_\Gamma,\alpha_\Lambda}(\mathbf{\Gamma}_{t+1},\mathbf{\Lambda}_{t+1},\mathbf{U},\mathbf{V}_t;\mathbf{A}_{\Gamma,t},\mathbf{A}_{\Lambda,t})$
4. $\mathbf{V}_{t+1}=\arg\min_{\mathbf{V}} f_{\alpha_\Gamma,\alpha_\Lambda}(\mathbf{\Gamma}_{t+1},\mathbf{\Lambda}_{t+1},\mathbf{U}_{t+1},\mathbf{V};\mathbf{A}_{\Gamma,t},\mathbf{A}_{\Lambda,t})$
5. $\mathbf{A}_{\Gamma,t+1}\doteq\mathbf{A}_{\Gamma,t}+\alpha_\Lambda(\mathbf{\Gamma}_{t+1}-\mathbf{U}_{t+1})$
6. $\mathbf{A}_{\Lambda,t+1}\doteq\mathbf{A}_{\Lambda,t}+\alpha_\Gamma(\mathbf{\Gamma}_{t+1}-\mathbf{V}_{t+1})$

1) Update step for $\mathbf{\Gamma}$: The first suboptimization problem involves the minimization of $f_{\alpha_\Gamma,\alpha_\Lambda}(\mathbf{\Gamma},\mathbf{\Lambda},\mathbf{V},\mathbf{U};\mathbf{A}_\Lambda,\mathbf{A}_\Gamma)$ with respect to $\mathbf{\Gamma}$. It has the quadratic structure, which is easy to solve by setting the first-order derivative equal to zero. Furthermore, the loss function $\mathcal{C}(\mathbf{\Gamma},\mathbf{\Lambda})$ is a sum of convex functions associated with sub-matrices $\mathbf{\Gamma}^i$, one can seek for $\mathbf{\Gamma}^i_{t+1},\ i=1,\ldots,D$, which has the following solution

$$\mathbf{\Gamma}^i_{t+1}=(\mathbf{X}^{i^T}\mathbf{X}^i+\alpha_\Gamma\mathbf{I})^{-1}(\mathbf{X}^{i^T}(\mathbf{Y}^i-\mathbf{\Lambda}^i_t)+\alpha_\Gamma\mathbf{V}^i_t+\mathbf{A}^i_{V,t}),$$

where $\mathbf{I}$ is $p\times p$ identity matrix and $\mathbf{\Lambda}^i_t,\mathbf{V}^i_t$ and $\mathbf{A}^i_{V,t}$ are submatrices of $\mathbf{\Lambda}_t,\mathbf{V}_t$ and $\mathbf{A}_{V,t}$, respectively.

2) Update step for $\mathbf{\Lambda}$: The second suboptimization problem is similar in nature, whose solution is given below

$$\mathbf{\Lambda}^i_{t+1}=(1+\alpha_\Lambda)^{-1}(\mathbf{Y}^i-\mathbf{X}^i\mathbf{\Gamma}^i_{t+1}+\alpha_\Lambda\mathbf{U}^i_t-\mathbf{A}^i_{\Lambda,t}),$$

where $\mathbf{U}^i_t$ and $\mathbf{A}^i_{\Lambda,t}$ are submatrices of $\mathbf{U}_t$ and $\mathbf{A}_{\Lambda,t}$, respectively.

3) Update step for $\mathbf{U}$: The third suboptimization problem is with respect to $\mathbf{U}$, which is the standard ℓ_1 minimization problem which can be recast as

$$\min_{\mathbf{U}}\frac{1}{2}\|\mathbf{\Lambda}_{t+1}+\alpha_\Lambda^{-1}\mathbf{A}_{\Lambda,t}-\mathbf{U}\|_F^2+\frac{\lambda_2}{\alpha_\Lambda}\|\mathbf{U}\|_1. \quad (10)$$

Equation (10) is the well-known shrinkage problem whose solution is given by

$$\mathbf{U}_{t+1}=\mathcal{S}\left(\mathbf{\Lambda}_{t+1}+\alpha_\Lambda^{-1}\mathbf{A}_{\Lambda,t},\frac{\lambda_2}{\alpha_\Lambda}\right),$$

where $\mathcal{S}(a,b)=sgn(a)(|a|-b)$ for $|a|\geq b$ and zero otherwise.

4) Update step for $\mathbf{V}$: The final suboptimization problem is with respect to $\mathbf{V}$ and can be reformulated as

$$\min_{\mathbf{V}}\frac{1}{2}\|\mathbf{\Gamma}_{t+1}+\alpha_\Gamma^{-1}\mathbf{A}_{\Gamma,t}-\mathbf{V}\|_F^2+\frac{\lambda_1}{\alpha_\Gamma}\|\mathbf{V}\|_{1,q}. \quad (11)$$

Due to the separable structure of (11), it can be solved by minimizing with respect to each row of $\mathbf{V}$ separately. Let $\boldsymbol{\gamma}_{i,t+1},\mathbf{a}_{\Gamma,i,t}$ and $\mathbf{v}_{i,t+1}$ be rows of matrices $\mathbf{\Gamma}_{t+1},\mathbf{A}_{\Gamma,t}$ and $\mathbf{V}_{t+1}$, respectively. Then for each $i=1,\ldots,p$ we solve the following sub-problem

$$\mathbf{v}_{i,t+1}=\arg\min_{\mathbf{v}}\frac{1}{2}\|\mathbf{z}-\mathbf{v}\|_2^2+\eta\|\mathbf{v}\|_q, \quad (12)$$

where $\mathbf{z}=\boldsymbol{\gamma}_{i,t+1}-\mathbf{a}_{\Gamma,i,t}\alpha_\Gamma^{-1}$ and $\eta=\frac{\lambda_1}{\lambda_2}$. One can derive the solution for (12) for any q. In this paper, we only focus on the case when $q=2$. The solution of (12) has the following form

$$\mathbf{v}_{i,t+1}=\left(1-\frac{\eta}{\|\mathbf{z}\|_2}\right)_+\mathbf{z},$$

where $(\mathbf{v})_+$ is a vector with entries receiving values $\max(v_i,0)$.

Our proposed algorithm Sparse Multimodal Biometrics Recognition (SMBR) is summarized in Algorithm 2. We refer to the robust method taking sparse error into account as SMBR-E (SMBR with error), and the initial case where it is not taken account as SMBR-WE (SMBR without error).

Algorithm 2: Sparse Multimodal Biometrics Recognition (SMBR).

Input: Training samples $\{\mathbf{X_i}\}_{i=1}^D$, test sample $\{\mathbf{Y_i}\}_{i=1}^D$, Occlusion basis $\{\mathbf{B}\}_{i=1}^D$
Procedure: Obtain $\hat{\mathbf{\Gamma}}$ and $\hat{\mathbf{\Lambda}}$ by solving

$$\hat{\mathbf{\Gamma}},\hat{\mathbf{\Lambda}}=\arg\min_{\mathbf{\Gamma},\mathbf{\Lambda}}\frac{1}{2}\sum_{i=1}^D\|\mathbf{Y}^i-\mathbf{X}^i\mathbf{\Gamma}^i-\mathbf{B}^i\mathbf{\Lambda}^i\|_F^2+\lambda_1\|\mathbf{\Gamma}\|_{1,q}+\lambda_2\|\mathbf{\Lambda}\|_1,$$

Output:
$\texttt{identity}(\mathbf{Y})=\arg\min_j\sum_{i=1}^D\|\mathbf{Y}^i-\mathbf{X}^i\boldsymbol{\delta}^i_j(\hat{\mathbf{\Gamma}}^i)-\mathbf{B}^i\hat{\mathbf{\Lambda}}^i\|_F^2.$

III. Quality based fusion

Ideally a fusion mechanism should give more weights to the more reliable modalities. Hence, the concept of quality is important in multimodal fusion. A quality measure based on sparse representation was introduced for faces in [10]. To decide whether a given test sample has good quality or not, its Sparsity Concentration Index (SCI) was calculated. Given a coefficient vector $\gamma\in\mathbb{R}^p$, the SCI is given as:

$$SCI(\boldsymbol{\gamma})=\frac{\frac{C.\max_{i\in\{1,\cdots,C\}}\|\delta_i(\boldsymbol{\gamma})\|_1}{\|\boldsymbol{\gamma}\|_1}-1}{C-1}$$

where, δ_i is the indicator function keeping the coefficients corresponding to the i^{th} class and setting others to zero. SCI values close to 1 correspond to the case where the test sample can be represented well using the samples of a single class, hence is of high quality. On the other hand, samples with SCI close to 0 are not similar to any of the classes, and hence are of poor quality. This can be easily extended to the multimodal

case using the joint sparse representation matrix $\hat{\mathbf{\Gamma}}$. In this case, we can define the quality, q_j^i for sample $\mathbf{y}_j^i$ as:

$$q_j^i = SCI(\hat{\mathbf{\Gamma}}^i_j)$$

where, $\hat{\mathbf{\Gamma}}^i_j$ is the j^{th} column of $\hat{\mathbf{\Gamma}}^i$. Given this quality measure, the classification rule (2) can be modified to include the quality measure.

$$\hat{j} = \arg\min_j \sum_{i=1}^{D}\sum_{k=1}^{d_i} q_k^i \|\mathbf{y}_k^i - \mathbf{X}^i\boldsymbol{\delta}_j(\mathbf{\Gamma}_k^i)\|_F^2, \qquad (13)$$

where, $\boldsymbol{\delta}_j$ is the indicator function retaining the coefficients corresponding to j^{th} class.

IV. Kernel space multimodal biometrics recognition

The class identities in the multibiometric dataset may not be linearly separable. Hence, we also extend the sparse multimodal fusion framework to kernel space. The kernel function, $\kappa : \mathbb{R}^n \times \mathbb{R}^n$, is defined as the inner product

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$$

where, ϕ is an implicit mapping projecting the vector $\mathbf{x}$ into a higher dimensional space.

A. Multivariate kernel sparse representation

Considering the general case of D modalities with $\{\mathbf{Y}^i\}_{i=1}^D$ as a set of d_i observations, the feature space representation can be written as:

$$\mathbf{\Phi}(\mathbf{Y}^i) = [\phi(\mathbf{y}_1^i), \phi(\mathbf{y}_2^i), ..., \phi(\mathbf{y}_d^i)]$$

Similarly, the dictionary of training samples for modality $i = 1, \cdots, D$ can be represented in feature space as

$$\mathbf{\Phi}(\mathbf{X}^i) = [\phi(\mathbf{X}_1^i), \phi(\mathbf{X}_2^i), \cdots, \phi(\mathbf{X}_C^i)]$$

As in joint linear space representation, we have:

$$\mathbf{\Phi}(\mathbf{Y}^i) = \mathbf{\Phi}(\mathbf{X}^i)\mathbf{\Gamma}^i$$

where, $\mathbf{\Gamma}^i$ is the coefficient matrix associated with modality i. Incorporating information from all the sensors, we seek to solve the following optimization problem similar to the linear case:

$$\hat{\mathbf{\Gamma}} = \arg\min_{\mathbf{\Gamma}} \frac{1}{2}\sum_{i=1}^{D} \|\mathbf{\Phi}(\mathbf{Y}^i) - \mathbf{\Phi}(\mathbf{X}^i)\mathbf{\Gamma}^i\|_F^2 + \lambda\|\mathbf{\Gamma}\|_{1,q} \quad (14)$$

where, $\mathbf{\Gamma} = [\mathbf{\Gamma}^1, \mathbf{\Gamma}^2, \cdots, \mathbf{\Gamma}^D]$. It is clear that the information from all modalities are integrated via the shared sparsity pattern of the matrices $\{\mathbf{\Gamma}^i\}_{i=1}^D$. This can be reformulated in terms of kernel matrices as:

$$\hat{\mathbf{\Gamma}} = \arg\min_{\mathbf{\Gamma}} \frac{1}{2}\sum_{i=1}^{D} \big(\text{trace}(\mathbf{\Gamma^i}^T \mathbf{K}_{\mathbf{X}_i,\mathbf{X}_i}\mathbf{\Gamma}^i) - 2\text{trace}(\mathbf{K}_{\mathbf{X}_i,\mathbf{Y}_i}\mathbf{\Gamma}^i)\big) + \lambda\|\mathbf{\Gamma}\|_{1,q} \qquad (15)$$

where, the kernel matrix $\mathbf{K}_{\mathbf{A},\mathbf{B}}$ is defined as:

$$\mathbf{K}_{\mathbf{A},\mathbf{B}}(i,j) = \langle \phi(\mathbf{a}_i), \phi(\mathbf{b}_j) \rangle \qquad (16)$$

$\mathbf{a}_i$ and $\mathbf{b}_j$ being i^{th} and j^{th} columns of $\mathbf{A}$ and $\mathbf{B}$ respectively.

B. Composite kernel sparse representation

Another way to combine information of different modalities is through composite kernel, which efficiently combines kernel for each modality. The kernel combines both within and between similarities of different modalities. For two modalities with the same feature dimension, the kernel matrix can be constructed as:

$$\kappa(\mathbf{X}_i, \mathbf{X}_j) = \alpha_1\kappa(\mathbf{x}_i^1, \mathbf{x}_j^1) + \alpha_2\kappa(\mathbf{x}_i^1, \mathbf{x}_j^2) + \alpha_3\kappa(\mathbf{x}_i^2, \mathbf{x}_j^1) + \alpha_4\kappa(\mathbf{x}_i^2, \mathbf{x}_j^2) \qquad (17)$$

where, $\{\alpha_i\}_{i=1,\cdots,4}$ are the weights associated with the kernels and $\mathbf{X}_i = [\mathbf{x}_i^1; \mathbf{x}_i^2]$. $\mathbf{x}_i^1$ and $\mathbf{x}_i^2$ are the feature vectors for modality 1 and 2 respecetively. It can be similarly extended to multiple modalities. However, the modalities may be of different dimensions. In such cases, cross-simlarity measure is not possible. Hence, the modalities are divided according to being homogenous (e.g. right and left iris) or heterogenous (fingerprint and iris), as homogenous modalities have same feature extraction process and hence, same dimension. This is also reasonable, because homogenous modalities are correlated at feature level but heteregenous modalities may not be correlated. A composite kernel is defined for each homogenous modality. For D modalities, with $\{d_i\}_{i\in\mathcal{S}_j}$, $\mathcal{S}_j \subseteq \{1, 2, \cdots, D\}$ being the sets of indices of each homogenous modality, the composite kernel for each set will be given as:

$$\kappa(\mathbf{X}_i^k, \mathbf{X}_j^k) = \sum_{s_1 s_2 \in \mathcal{S}_k} \alpha_{s_1 s_2}\kappa(\mathbf{x}_i^{s_1}, \mathbf{x}_j^{s_2}) \qquad (18)$$

Here, $\mathbf{X}_i^k = [\mathbf{x}_i^{s_1}; \mathbf{x}_i^{s_2}; \cdots; \mathbf{x}_i^{s_{|\mathcal{S}_k|}}]$, $\mathcal{S}_k = [s_1, s_2, \cdots, s_{|\mathcal{S}_k|}]$ and $k = 1, \cdots, N_H$, N_H being the number of different heterogenous modalities. The information from the different heterogenous modalities can then be combined similar to the sparse kernel fusion case:

$$\hat{\mathbf{\Gamma}} = \arg\min_{\mathbf{\Gamma}} \frac{1}{2}\sum_{i=1}^{N_H} \big(\text{trace}(\mathbf{\Gamma^i}^T \mathbf{K}_{\mathbf{X}^i,\mathbf{X}^i}\mathbf{\Gamma}^i) - 2\text{trace}(\mathbf{K}_{\mathbf{X}^i,\mathbf{Y}^i}\mathbf{\Gamma}^i)\big) + \lambda\|\mathbf{\Gamma}\|_{1,q} \qquad (19)$$

where, $\mathbf{K}_{\mathbf{X}^i,\mathbf{X}^i}$ is defined for each $\mathcal{S}_i$ as in (16) and $\mathbf{\Gamma} = [\mathbf{\Gamma}^1, \mathbf{\Gamma}^2, \cdots, \mathbf{\Gamma}^{N_H}]$.

C. Optimization Algorithm

Similar to the linear fusion method, we apply the alternating direction method to efficiently solve the problem for kernel fusion. The method splits the variable $\mathbf{\Gamma}$ such that the new problem has two convex functions. This is done by introducing a new variable $\mathbf{V}$ and reformulating the problems (15) and (19) as:

$$\arg\min_{\mathbf{\Gamma},\mathbf{V}} \frac{1}{2}\sum_{i=1}^{N_K} \big(\text{trace}(\mathbf{\Gamma^i}^T \mathbf{K}_{\mathbf{X}^i,\mathbf{X}^i}\mathbf{\Gamma}^i) - 2\text{trace}(\mathbf{K}_{\mathbf{X}^i,\mathbf{Y}^i}\mathbf{\Gamma}^i)\big) + \lambda\|\mathbf{V}\|_{1,q} \text{s.t.} \mathbf{\Gamma} = \mathbf{V} \qquad (20)$$

where, N_K is the number of kernels in (15) and (19). Rewriting the problem using Lagrangian multiplier, the optimization

becomes:

$$\arg\min_{\mathbf{\Gamma},\mathbf{V}} \frac{1}{2}\sum_{i=1}^{N_K}\big(\text{trace}({\mathbf{\Gamma}^i}^T\mathbf{K}_{\mathbf{X}^i,\mathbf{X}^i}\mathbf{\Gamma}^i) - 2\text{trace}(\mathbf{K}_{\mathbf{X}^i,\mathbf{Y}^i}\mathbf{\Gamma}^i)\big) + \lambda\|\mathbf{V}\|_{1,q} + \langle\mathbf{B},\mathbf{\Gamma}-\mathbf{V}\rangle + \frac{\beta_W}{2}\big\|\mathbf{\Gamma}-\mathbf{V}\big\|_F^2 \quad (21)$$

which upon re-arranging becomes:

$$\arg\min_{\mathbf{\Gamma},\mathbf{V}} \frac{1}{2}\sum_{i=1}^{N_K}\big(\text{trace}({\mathbf{\Gamma}^i}^T\mathbf{K}_{\mathbf{X}^i,\mathbf{X}^i}\mathbf{\Gamma}^i) - 2\text{trace}(\mathbf{K}_{\mathbf{X}^i,\mathbf{Y}^i}\mathbf{\Gamma}^i)\big) + \lambda\|\mathbf{V}\|_{1,q} + \frac{\beta_W}{2}\big\|\mathbf{\Gamma}-\mathbf{V}+\frac{1}{\beta_W}\mathbf{B}\big\|_F^2 \quad (22)$$

The optimization algorithm is summarized in Algorithm 3. Each of the optimization steps has simple closed-form expression.

Algorithm 3: Alternating Direction Method of Multipliers (ADMM) in kernel space.

Initialize: $\mathbf{\Gamma}_0, \mathbf{V}_0, \mathbf{B}_0, \beta_W$
While not converged do
1. $\mathbf{\Gamma}_{t+1} = \arg\min_{\mathbf{\Gamma}} \frac{1}{2}\sum_{i=1}^{N_K}\big(\text{trace}({\mathbf{\Gamma}^i}^T\mathbf{K}_{\mathbf{X}^i,\mathbf{X}^i}\mathbf{\Gamma}^i) - 2\text{trace}(\mathbf{K}_{\mathbf{X}^i,\mathbf{Y}^i}\mathbf{\Gamma}^i)\big) + \lambda\|\mathbf{V}_t\|_{1,q} + \frac{\beta_W}{2}\big\|\mathbf{\Gamma}-\mathbf{V_t}+\frac{1}{\beta_W}\mathbf{B_t}\big\|_F^2$
2. $\mathbf{V}_{t+1} = \arg\min_{\mathbf{V}} \lambda\|\mathbf{V}\|_{1,q} + \frac{\beta_W}{2}\big\|\mathbf{\Gamma}_{t+1}-\mathbf{V_t}+\frac{1}{\beta_W}\mathbf{B_t}\big\|_F^2$
3. $\mathbf{B}_{t+1} = \mathbf{B}_t + \beta_W(\mathbf{\Gamma}_{t+1}-\mathbf{V}_{t+1})$

1) Update steps for $\mathbf{\Gamma}_t$: $\mathbf{\Gamma}_{t+1}$ is obtained by updating each submatrix $\mathbf{\Gamma}_t^i$, $i = 1, \cdots, N_K$ as:

$$\mathbf{\Gamma}_t^i = (\mathbf{K}_{\mathbf{X}^i,\mathbf{X}^i} + \beta_W\mathbf{I})^{-1}(\mathbf{K}_{\mathbf{X}^i,\mathbf{Y}^i} + \beta_W\mathbf{V}_t^i - \mathbf{B}_t^i) \quad (23)$$

where, $\mathbf{I}$ is an identity matrix and $\mathbf{V}_t^i$, $\mathbf{B}_t^i$ are sub-matrices of $\mathbf{V}_t$ and $\mathbf{B}_t$ respectively.

2) Update steps for $\mathbf{V}_t$: The update equation for $\mathbf{V}_t$ is same as in the linear fusion case using (11) and (12), replacing $\mathbf{A}_{\Gamma,t}$ and α_Γ with $\mathbf{B}_t$ and β_W respectively.

D. Classification

Once $\mathbf{\Gamma}$ is obtained using any of the two methods above, classification can be done by assigning the class label as:

$$\hat{j} = \arg\min_j \sum_{i=1}^{N_K}\|\mathbf{\Phi}(\mathbf{Y}^i) - \mathbf{\Phi}(\mathbf{X}_j^i)\hat{\mathbf{\Gamma}}_j^i\|_F^2$$

or in terms of kernel matrices as:

$$\hat{j} = \arg\min_j \sum_{i=1}^{N_K}\big(\text{trace}(\mathbf{K}_{\mathbf{YY}}) - 2\text{trace}(\hat{\mathbf{\Gamma}}_j^{i^T}\mathbf{K}_{\mathbf{X_j^i Y}}\hat{\mathbf{\Gamma}}_j^i) + \text{trace}(\hat{\mathbf{\Gamma}}_j^{i^T}\mathbf{K}_{\mathbf{X_j^i X_j^i}}\hat{\mathbf{\Gamma}}_j^i)\big) \quad (24)$$

Here, $\mathbf{X}_j^i$ is the sub-dictionary associated with j^{th} class and $\hat{\mathbf{\Gamma}}_j^i$ is the coefficient matrix associated with this class.

The classification rule can be further extended to include the quality measure as in (13). But, we skip this step here, as we wish to study the effect of kernel representation and quality separately.

Multivariate Kernel Sparse Recognition (kerSMBR) and Composite Kernel Sparse Recognition (compSMBR) algorithms are summarized in Algorithms 4 and 5, respectively:

Algorithm 4: Kernel Sparse Multimodal Biometrics Recognition (kerSMBR).

Input: Training samples $\{\mathbf{X_i}\}_{i=1}^D$, test sample $\{\mathbf{Y_i}\}_{i=1}^D$
Procedure: Obtain $\hat{\mathbf{\Gamma}}$ by solving

$$\hat{\mathbf{\Gamma}} = \arg\min_{\mathbf{\Gamma}} \frac{1}{2}\sum_{i=1}^{D}\|\mathbf{\Phi}(\mathbf{Y}^i) - \mathbf{\Phi}(\mathbf{X}^i)\mathbf{\Gamma}^i\|_F^2 + \lambda\|\mathbf{\Gamma}\|_{1,q} \quad (25)$$

Output: $\texttt{identity}(\mathbf{Y}) = \arg\min_j \sum_{i=1}^D \big(\text{trace}(\mathbf{K}_{\mathbf{YY}}) - 2\text{trace}(\hat{\mathbf{\Gamma}}_j^{i^T}\mathbf{K}_{\mathbf{X_j^i Y}}\hat{\mathbf{\Gamma}}_j^i) + \text{trace}(\hat{\mathbf{\Gamma}}_j^{i^T}\mathbf{K}_{\mathbf{X_j^i X_j^i}}\hat{\mathbf{\Gamma}}_j^i)\big)$

Algorithm 5: Composite Kernel Sparse Multimodal Biometrics Recognition (compSMBR).

Input: Training samples $\{\mathbf{X_i}\}_{i=1}^D$, test sample $\{\mathbf{Y_i}\}_{i=1}^D$
Procedure: Obtain $\hat{\mathbf{\Gamma}}$ by solving

$$\hat{\mathbf{\Gamma}} = \arg\min_{\mathbf{\Gamma}} \frac{1}{2}\sum_{i=1}^{N_H}\big(\text{trace}({\mathbf{\Gamma}^i}^T\mathbf{K}_{\mathbf{X}^i,\mathbf{X}^i}\mathbf{\Gamma}^i) - 2\text{trace}(\mathbf{K}_{\mathbf{X}^i,\mathbf{Y}^i}\mathbf{\Gamma}^i)\big) + \lambda\|\mathbf{\Gamma}\|_{1,q} \quad (26)$$

Output: $\texttt{identity}(\mathbf{Y}) = \arg\min_j \sum_{i=1}^{N_H} \big(\text{trace}(\mathbf{K}_{\mathbf{YY}}) - 2\text{trace}(\hat{\mathbf{\Gamma}}_j^{i^T}\mathbf{K}_{\mathbf{X_j^i Y}}\hat{\mathbf{\Gamma}}_j^i) + \text{trace}(\hat{\mathbf{\Gamma}}_j^{i^T}\mathbf{K}_{\mathbf{X_j^i X_j^i}}\hat{\mathbf{\Gamma}}_j^i)\big)$

V. Experiments

We evaluated our algorithm for different multi-biometric settings. We tested on two publicly available datasets - the WVU Multimodal dataset [27] and the AR face dataset [28]. The WVU dataset is one of the few publicly available datasets which allows fusion at image level. It is a challenging dataset having samples from differnent biometric modalities for each subject.

In the second experiment, we show the applicability of our method to fusing information from *soft* biometrics. Recently, combining information from soft biometrics such as facial marks, hair color, etc has been shown to improve face recognition performance [29]. The challenge for fusion algorithms is combine these weak modalities with strong modalities as face or fingerprint [30]. We demonstrate that our framework can be extended to address this problem. Further, we also show the effect of noise and occlusion on performance of different algorithms. In all the experiments $\mathbf{B}_i$ was set to be identity for convenience, *i.e.*, we assume noise to be sparse in image domain.

A. WVU Multimodal Dataset

The WVU multimodal dataset is a comprehensive collection of different biometric modalities such as fingerprint, iris, palmprint, hand geometry and voice from subjects of different age, gender and ethnicity as described in Table I. It is a challenging dataset and many of these samples are corrupted with blur, occlusion and sensor noise as shown in Figure 2. Out of these, we chose iris and fingerprint modalities for testing the proposed algorithms. In total, there are 2 iris (right and left iris) and 4 fingerprint modalities. Also, the evaluation was done on a subset of 219 subjects having samples in both modalities.

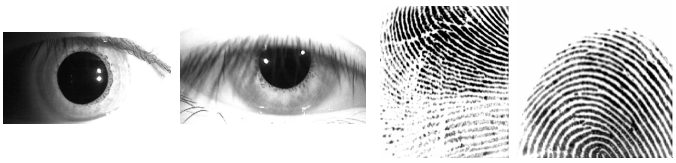

Fig. 2: Examples of challenging images from the WVU Multimodal dataset. The images above suffer from various artifacts as sensor noise, blur, occlusion and poor acquisition.

Biometric Modality	# of subjects	# of samples
Iris	244	3099
Fingerprint	272	7219
Palm	263	683
Hand	217	3062
Voice	274	714

TABLE I: WVU Biometric Data

1) Preprocessing: Robust pre-processing of images was done before feature extraction. Iris images were segmented following the recent method proposed in [31]. Following the segmentation, 25×240 iris templates were formed by re-sampling using the publicly available code of Masek *et al.* [32]. Fingerprint images were enhanced using the filtering based methods described in [33], and then the core point was detected using the enhanced images [34]. Features were then extracted around the detected core point.

2) Feature Extraction: Gabor features were extracted on the processed images as they have been shown to give good performance on both fingerprints [34] and iris [35]. For fingerprint samples, the processed images were convolved with Gabor filters at 8 different orientations. Circular tesselations were extracted around the core point for all the filtered images similar to [34]. The tesselation consisted of 15 concentric bands, each of width 5 pixels and divided into 30 sectors. The mean values for each sector were concatenated to form the feature vector of size 3600×1. Features for iris images were formed by convolving the templates with log-Gabor filter at a single scale, and vectorizing the template to give a 6000×1 dimensional feature.

3) Experimental Set-up: The dataset was randomly divided into 4 training samples per class (1 sample here is 1 data sample each from 6 modalities) and the rest 519 for testing. The recognition result was averaged over 5 runs. The proposed methods were compared with state-of-the-art classification methods such as sparse logistic regression (SLR) [36] and SVM [37]. Although these methods have been shown to give superior performance, they cannot handle multimodal data. One possible way to handle multimodal data is to use feature concatenation. But, this resulted in feature vectors of size 26400×1 when all six modalities are used, and is not useful. Hence, we explored score-level and decision-level fusion methods for combining results of individual modalities. For score-level fusion, the probability outputs for test sample of each modality, $\{\mathbf{y}_i\}_{i=1}^{6}$ were added together to give the final score vector. Classification was based upon the final score values. For decision-level fusion, the subject chosen by the maximum number of modalities was taken to be from the correct class. We tested the proposed linear and kernel fusion techniques separately and compared them with the linear and kernel versions of SLR and SVM respectively. We denote the score-level fusion of these methods as SLR-Sum and SVM-Sum, and the decision-level fusion as SLR-Major and SVM-Major.

*a) **Linear Fusion**:* The recognition performances of SMBR-WE and SMBR-E was compared with linear SVM and linear SLR classification methods. The parameter values λ_1 and λ_2 were set to 0.01.

- *Comparsion of Methods:* Figures 3 shows the performance on individual modalities. All the classifiers show similar trend. The performance for all of them are lower on iris images and fingers 1 and 3. However, the SVM fares poorer than other methods on all the modalities. Figure 4 and Table II show performance for different fusion settings. The proposed SMBR approach outperforms existing classification techniques. Both SMBR-E and SMBR-WE have similar performance, though the latter seems to give a slightly better performance. This may be due to the penalty on the sparse error, though the error may not be sparse in the image domain. Further, sum-based fusion shows a superior performance over voting-based methods.
- *Fusion with quality:* Clearly different modalities have different quality of performance. Hence, we studied the effect of the proposed quality measure on the performance of different methods. For a consistent comparison, the quality values produced by SMBR-E method was used for all the algorithms. Table III shows the performance for the three fusion settings. The effect of including the quality measure while classification can be studied by comparing with Table II. Clearly, the recognition rate increases for all the methods across the fusion settings. Again SMBR-E and SMBR-WE give the best performances among all the methods.
- *Effect of joint sparsity:* We also studied the effect of joint sparsity constraint on the recognition performance. For this, SMBR-WE algorithm was run for different values of λ_1. Figure 5 shows the rank one recognition variation across λ_1 values for different fusion settings. All the curves show a sharp increase in performance around $\lambda_1 = 0$. Further, the increase is more for iris fusion, which shows around 5% improvement at $\lambda_1 = 0.005$ over $\lambda_1 = 0$. This shows that imposing joint sparsity constraint is important for fusion. Moreover, it helps in regulating fusion performance, when reconstruction error alone is not sufficient to distinguish between different classes. The performance is then stable across λ_1 values, and starts decreasing slowly after reaching the optimum performance.
- *Variation with number of training samples:* We varied the number of training samples and studied the effect on the top 4 algorithms. Figure 6 shows the variation for fusion of all the modalities. It can be seen that SMBR-WE and SMBR-E are stable across number of training samples, whereas the performance of SLR and SVM based methods fall sharply. The fall in performance of

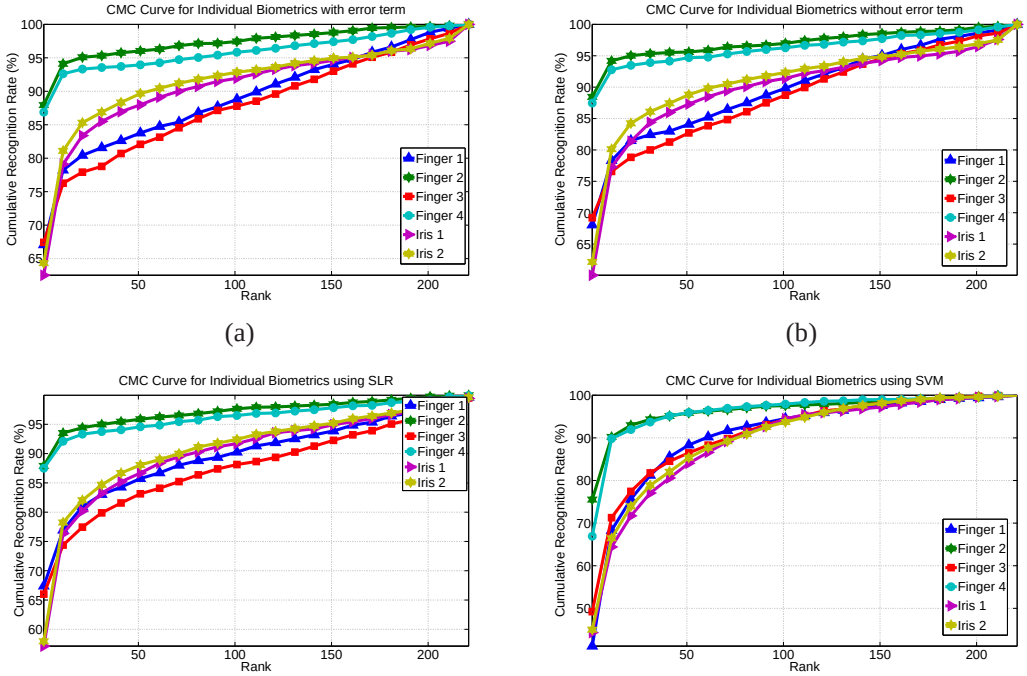

Fig. 3: CMCs (Cumulative Match Curve) for individual modalities using (a) SMBR-E, (b) SMBR-WE, (c) SLR and (d) SVM methods.

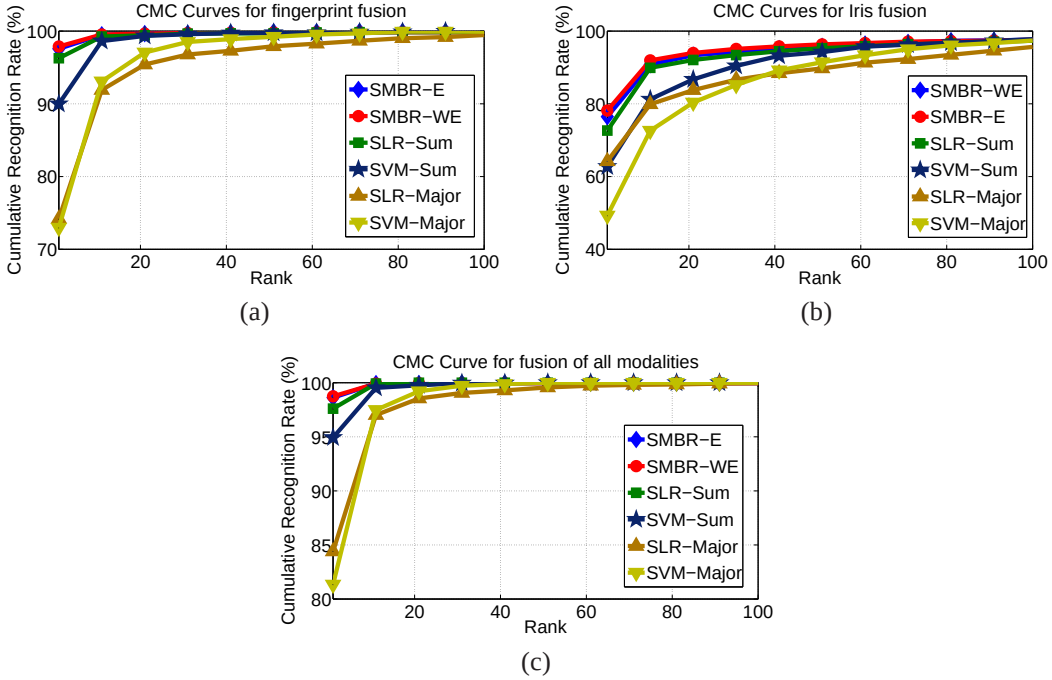

Fig. 4: CMCs for multimodal fusion using (a) four fingerprints, (b) two irises and (c) all modalities.

	SMBR-WE	SMBR-E	SLR-Sum	SLR-Major	SVM-Sum	SVM-Major
4 Fingerprints	**97.9**	97.6	96.3	74.2	90.0	73.0
2 Irises	76.5	**78.2**	72.7	64.2	62.8	49.3
All modalities	**98.7**	98.6	97.6	84.4	94.9	81.3

TABLE II: Rank one recognition performance for the WVU Multimodal dataset.

	SMBR-WE	SMBR-E	SLR-Sum	SLR-Major	SVM-Sum	SVM-Major
4 Fingerprints	**98.2**	98.1	97.5	86.3	93.6	85.5
2 Irises	76.9	**78.8**	74.1	67.2	64.3	51.6
All modalities	**98.8**	98.6	98.2	93.8	95.5	93.3

TABLE III: Rank one recognition performance using the proposed quality measure.

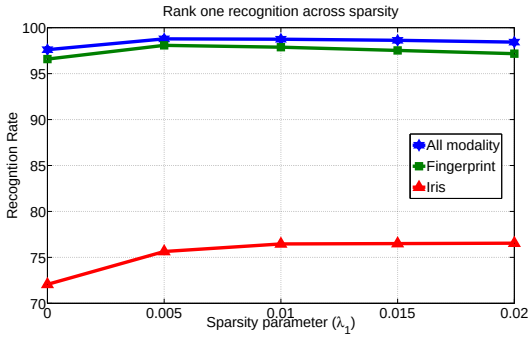

Fig. 5: Variation of recognition performance with different values of sparsity constraint, λ_1.

SLR and SVM can be attributed to the discriminative approaches of these methods, as well as score based fusion, as the fusion further reduces recognition when individual classifiers are not good.

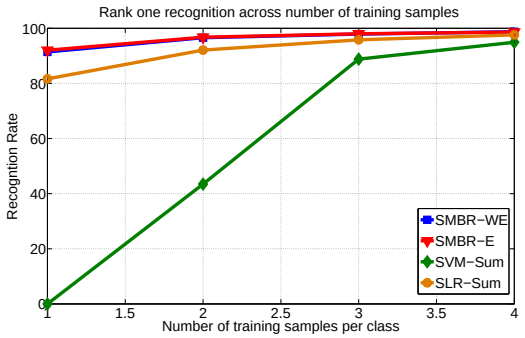

Fig. 6: Variation of recognition performance with number of training samples.

- *Comparison with other score-based fusion methods:* Although sum-based fusion is a popular technique for score fusion, some other techniques have also been proposed. We evaluated the performance of likelihood-based fusion method proposed in [38]. The results are shown in Table IV. The method does not show good performance as it models score distribution as Gaussian Mixture Model. However, it is difficult to model score distribution due to large variations in data samples. The method is also affected by the curse of dimensionality.

	2 irises	4 fingerprints	All modalities
SLR-Likelihood	66.6	83.5	75.1
SVM-Likelihood	50.7	31.9	31.0

TABLE IV: Fusion performance with likelihood-based method [38].

*b) **Kernel Fusion**:* We further compared the performances of proposed kerSMBR and compSMBR with kernel SVM and kernel SLR methods. In the experiments, we used Radial Basis Function (RBF) as the kernel, given as:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\sigma^2}\right),$$

σ being a parameter to control the width of the RBF.

- *Hyperparameter tuning:* To fix the value of hyperparameter, σ, we iterated over different values of σ, $\{2^{-3}, 2^{-2}, \cdots, 2^3\}$ for one set of training and test split of the data. The value of σ giving the maximum performance was fixed for each modality, and the performance was averaged over a few iterations. The weights $\{\alpha_{ij}\}$ were set to 1 for composite kernel. λ and β_W were set to 0.01 and 0.01 respectively.
- *Comparison of methods:* Figure 7 shows the performance of different methods on individual modalities, and Figure 8 and Table V on different fusion settings. Comparison of performance with linear fusion shows that the proposed kerSMBR significantly improves the performance on individual iris modalities as well as iris fusion. The performance on fingerprint modalities are similar, however the fusion of all 6 modalities shows an improvement of 0.4%. kerSMBR also achieves the best accuracy among all the methods for different fusion settings. kerSLR scores better than kerSVM in all the cases, and it's accuracy is close to kerSMBR. The performance of kerSLR is better than the linear counterpart, however kerSVM does not show much improvement. *Composite kernels* present an interesting case. Here, compSLR shows better performance than compSMBR on each homogenous modalities. Composite kernel by combining homogenous modalities into one, reduces effective number of modalities, hence the size of $\mathbf{\Gamma}$ matrix is reduced. This decreases the flexiblity in exploiting different modality information via $\mathbf{\Gamma}$. Hence, the performance of compSMBR is not optimal. It should be noted, however, we are not comparing composite kernels with kerSMBR as we have not optimized over α values for composite kernels. This is also a major limitation for composite kernel based methods.

B. AR Face Dataset

The AR face dataset consists of faces with varying illumination, expression and occlusion conditions, captured in two sessions. We evaluated our algorithms on a set of 100 users. Images from the first session, 7 for each subject were used as training and the images from the second session, again 7 per subject, were used for testing. Simple intensity values

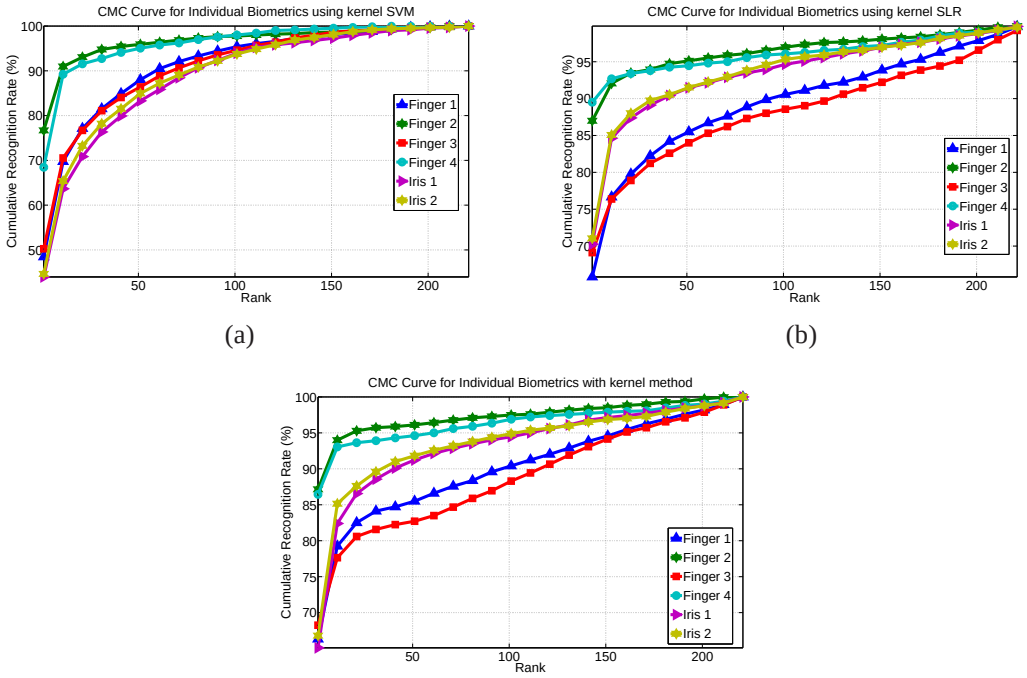

Fig. 7: CMCs for individual modalities using (a) kernel SVM, (b) kernel SLR and (c) kerSMBR.

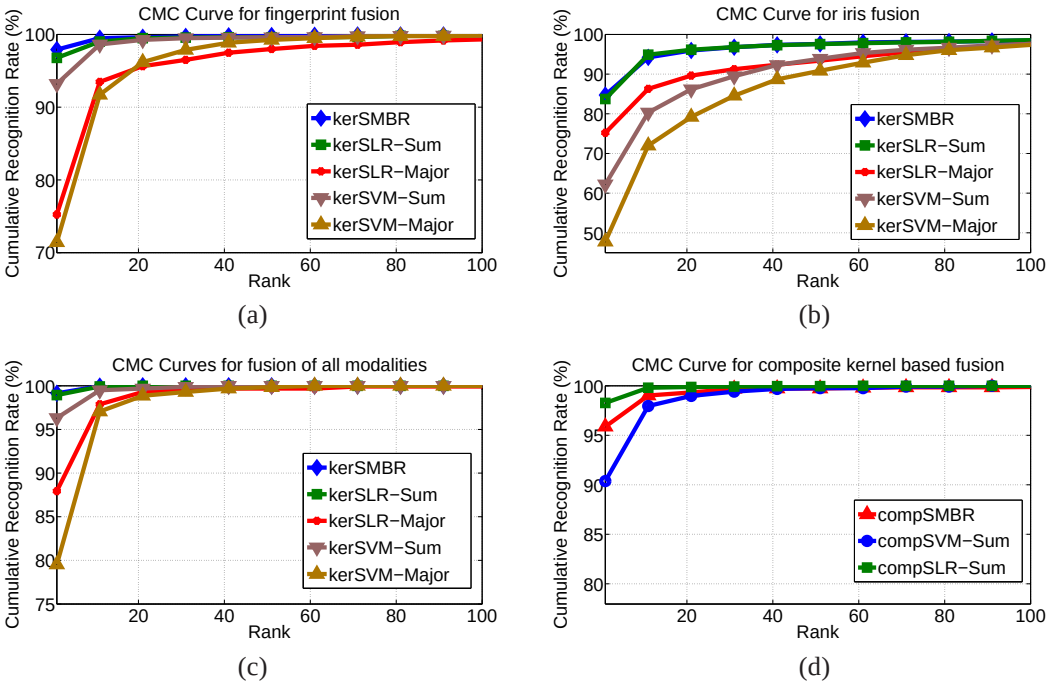

Fig. 8: CMCs for different fusion methods for (a) four fingerprints, (b) two irises and (c) all modalities. Results for composite kernels using different techniques is shown in figure (d).

	kerSMBR	kerSLR-Sum	kerSLR-Major	kerSVM-Sum	kerSVM-Major	compSMBR	compSLR-Sum	compSVM-Sum
4 Fingerprints	**97.9**	96.8	75.3	93.2	71.4	93.4	95.7	81.7
2 Irises	**84.7**	83.8	75.2	62.2	47.8	78.9	78.9	55.8
All modalities	**99.1**	98.9	87.9	96.3	79.5	95.9	98.2	90.4

TABLE V: Rank one recognition performance for the WVU Multimodal dataset.

were used as features. For each face, masks were applied around eye, mouth and nose regions, as shown in Figure 9, in order to provide four weak modalities. These, along with the whole face, were taken for fusion. The experimental set-up was similar to the previous section. The parameter values, λ_1 and λ_2 were set to 0.003 and 0.002 respectively. Furthermore, we also studied the effect of noise and occlusion on recognition performance.

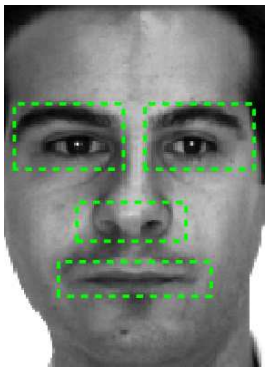

Fig. 9: Face mask used to crop out different modalities.

- *Comparison of methods:* Table VI shows the performance of different algorithms on the face dataset. Clearly, SMBR approach achieves about 5 % improvement over other techniques. Here, SR (sparse representation) shows the classification result using just the whole face. Evaluations for kernel techniques are not shown as linear kernel was found to be performing the best. FDDL [39] is a state-of-the-art discriminative dictionaries based technique, but using only single modality. Thus, by robustly classifying over multiple modalities, we achieve a remarkable improvement over the current benchmark. Further, a comparison with discriminative methods as SLR and SVM shows that they perform poorly compared to the proposed method. This is because weak modalities are hard to discriminate, hence score-level fusion with strong modality does not improve performance. On the other hand, by imposing reconstruction and joint sparsity simultaneously, the proposed method is able to achieve superior performance.
- *Effect of noise:* In this experiment, test images were corrupted with white Gaussian noise of increasing variance, σ^2. The comparisons are shown in Figure 10. It can be seen that both SMBR and SR methods are stable with noise.
- *Effect of occlusion:* In this experiment, a randomly chosen block of the test image was occluded. The recognition performance was studied with increasing block size. Figure 11 shows the performance of various algorithms with block size. SMBR-E is the most stable among all the methods due to robust handling of error. Recognition rates for others fall down sharply with increasing block size.
- *Quality based fusion:* Quality determination is an important parameter in fusion here, as a strong modality is being combined with weak modalities. We studied the effect of quality measure introduced in Section III. However, in this case we fix the quality for strong modality, *viz.* whole face to be 1, while for the weak modalities, the SCI values were taken. The recognition performance for SMBR-E and SMBR-WE across different noise and occlusion levels was studied. Figure 12 show the performance comparison with the unweighted methods. Using quality, the recognition performance for SMBR-WE goes up to 97.4 % from 96.9 %, whereas for SMBR-WE it increases to 97 % from 96 %. Similarly, results improve across different noise levels for both methods. However, SMBR-WE with quality shows worse performance as block size is increased. This may be because it does not handle sparse error, hence the quality values are not robust.

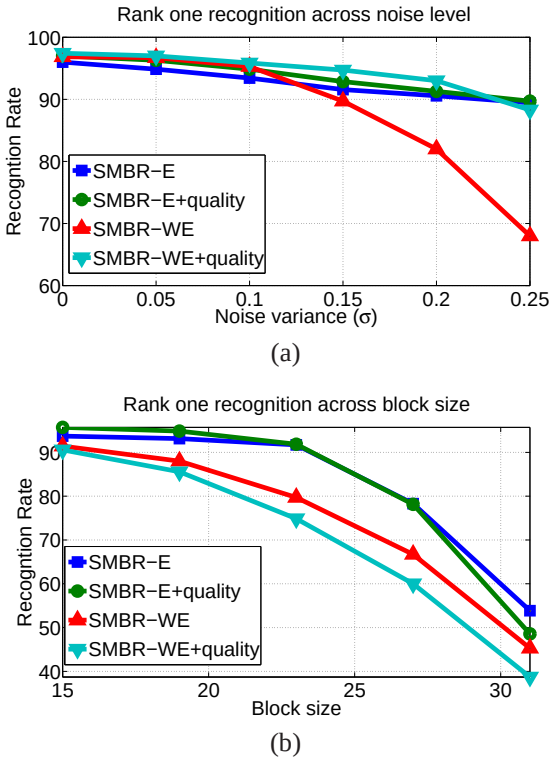

Fig. 12: Effect of quality on recognition performance across (a) noise (b) random blocks.

VI. Computational Complexity

The proposed algorithms are computationally efficient. The main steps of the algorithms are the update steps for $\mathbf{\Gamma}$, $\mathbf{\Lambda}$, $\mathbf{U}$ and $\mathbf{V}$. For linear fusion, the update step for $\mathbf{\Gamma}$ involves computing $(\mathbf{X}^{i^T}\mathbf{X}^i + \alpha_\Gamma \mathbf{I})^{-1}$ and four matrix multiplications. The first term is constant across iterations and can be precomputed. Matrix multiplication for two matrices of sizes $m \times n$ and $n \times p$ can be done in $\mathcal{O}(mnp)$ time. Hence, for given the training and test data, the computations are linear in feature dimension. Hence, large feature dimensions can be efficiently handled. Similarly, update step for $\mathbf{\Lambda}$ involves matrix multiplication $\mathbf{X}^i\mathbf{\Gamma}^i$. Update steps for $\mathbf{U}$ and $\mathbf{V}$ involves only scalar matrix computations and are very fast. Similarly in the kernel fusion, update for $\mathbf{\Gamma}$ involves calculating $(\mathbf{K}_{\mathbf{X}^i,\mathbf{X}^i} + \beta_W \mathbf{I})^{-1}$,

Method	Recognition Rate (%)	Method	Recognition Rate (%)
SMBR-WE	**96.9**	Linear SVM-Sum	86.7
SMBR-E	96	SLR-Sum	77.9
SR	91	FDDL [39]	91.9

TABLE VI: Rank one performance comparison of the proposed method.

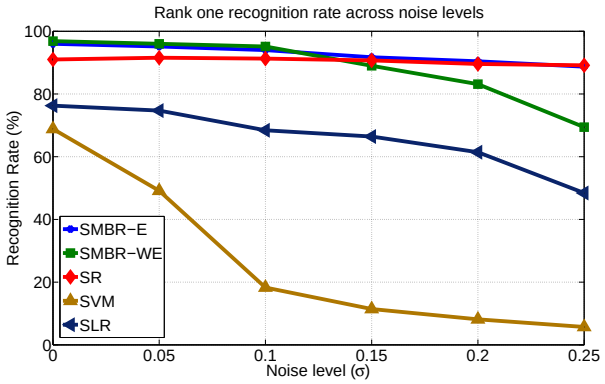

Fig. 10: Effect of noise on recognition performance.

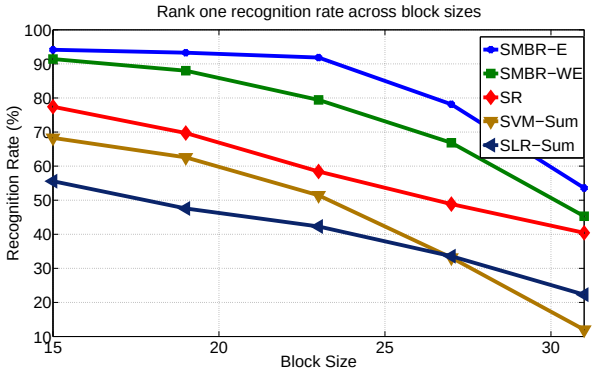

Fig. 11: Effect of occlusion on recognition performance.

which can be pre-computed. Other steps are similar to linear fusion. Classification step involves calculating the residual error for each class, and is efficient.

VII. Conclusion

We have proposed a novel joint sparsity-based feature level fusion algorithm for multimodal biometrics recognition. The algorithm is robust as it explicitly includes both noise and occlusion terms. An efficient algorithm based on alternative direction was proposed for solving the optimization problem. We also proposed a multimodal quality measure based on sparse representation. Further, the algorithm was extended to handle non-linear variations through kernel. Various experiments have shown that our method is robust and significantly improves the overall recognition accuracy.

Acknowledgment

The work of SS, VMP, and RC was partially supported by a MURI grant from the Army Research Office under the Grant W911NF-09-1-0383.

References

[1] A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*. Springer, 2006.

[2] A. Ross and A. K. Jain, "Multimodal biometrics: an overview," in *Proc. European Signal Processing Conference*, sept. 2004, pp. 1221–1224.

[3] P. Krishnasamy, S. Belongie, and D. Kriegman, "Wet fingerprint recognition: Challenges and opportunities," *International Joint Conference on Biometrics*, pp. 1–7, 2011.

[4] A. Klausner, A. Tengg, and B. Rinner, "Vehicle classification on multi-sensor smart cameras using feature- and decision-fusion," in *IEEE Conf. Dist. Smart Cameras*, Sept. 2007, pp. 67 –74.

[5] A. Rattani, D. Kisku, M. Bicego, and M. Tistarelli, "Feature level fusion of face and fingerprint biometrics," in *IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems*, Sept. 2007, pp. 1–6.

[6] X. Zhou and B. Bhanu, "Feature fusion of face and gait for human recognition at a distance in video," in *International Conference on Pattern Recognition*, vol. 4, 2006, pp. 529–532.

[7] A. A. Ross and R. Govindarajan, "Feature level fusion of hand and face biometrics," in *Proc. of the SPIE*, vol. 5779, 2005, pp. 196–204.

[8] V. M. Patel and R. Chellappa, "Sparse representations, compressive sensing and dictionaries for pattern recognition," in *Asian Conference on Pattern Recognition*, 2010.

[9] V. M. Patel, R. Chellappa, and M. Tistarelli, "Sparse representations and random projections for robust and cancelable biometrics," in *International Conference on Control, Automation, Robotics and Vision*, Dec. 2010, pp. 1–6.

[10] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern*

Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[11] J. K. Pillai, V. M. Patel, R. Chellappa, and N. K. Ratha, "Secure and robust iris recognition using random projections and sparse representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1877–1893, Sept. 2011.

[12] P. Nagesh and B. Li, "A compressive sensing approach for expression-invariant face recognition," *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1518–1525, Miami, FL, June 2009.

[13] V. M. Patel, T. Wu, S. Biswas, P. Phillips, and R. Chellappa, "Dictionary-based face recognition under variable lighting and pose," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 954 –965, June 2012.

[14] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2691–2698, 2010.

[15] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031 –1044, June 2010.

[16] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Towards a practical face recognition system: Robust alignment and illumination via sparse representation," To appear in *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2011, preprint: http://www.columbia.edu/ jw2966.

[17] S. Shekhar, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Joint sparsity-based robust multimodal biometrics recognition," in *European Conference on Computer Vision (ECCV) Workshop on Information Fusion in Computer Vision for Concept Recognition (IFCVCR)*, 2012.

[18] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B*, vol. 68, no. 1, pp. 49–67, 2006.

[19] L. Meier, S. V. D. Geer, and P. Bhlmann, "The group lasso for logistic regression," *Journal of the Royal Statistical Society: Series B*, vol. 70, no. 1, pp. 53–71, 2008.

[20] X.-T. Yuan and S. Yan, "Visual classification with multi-task joint sparse representation," in *International Conference on Computer Vision*, 2010.

[21] H. Zhang, N. M. Nasrabadi, Y. Zhang, and T. S. Huang, "Multi-observation visual recognition via joint dynamic sparse representation," in *International Conference on Computer Vision*, Nov. 2011.

[22] N. H. Nguyen, N. M. Nasrabadi, and T. D. Tran, "Robust multi-sensor classification via joint sparse representation," in *International Conference on Information Fusion*, 2011.

[23] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[24] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of ACM*, vol. 58, no. 1, pp. 1–37, 2009.

[25] J. Yang and Y. Zhang, "Alternating direction algorithms for l1 problems in compressive sensing," *SIAM Journal on Scientific Computing*, vol. 33, no. 1, pp. 250–278, 2011.

[26] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 681–695, 2011.

[27] S. S. S. Crihalmeanu, A. Ross and L. Hornak, "A protocol for multibiometric data acquisition, storage and dissemination," *In Technical Report, WVU, Lane Department of Computer Science and Electrical Engineering*, 2007.

[28] A. Martnez and R. Benavente, "The AR face database," *CVC Technical Report*, June 1998.

[29] U. Park and A. Jain, "Face matching and retrieval using soft biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 406–415, Sept. 2010.

[30] H. Li, K.-A. Toh, and L. Li, *Advanced Topics In Biometrics*. World Scientific Publishing Co. Pte. Ltd., 2012.

[31] S. Pundlik, D. Woodard, and S. Birchfield, "Non-ideal iris segmentation using graph cuts," in *IEEE CVPR Workshop*, June 2008, pp. 1–6.

[32] L. Masek and P. Kovesi, "MATLAB source code for biometric identification system based on iris patterns," The University of Western Australia, Tech. Rep., 2003.

[33] C. W. S. Chikkerur and V. Govindaraju, "A systematic approach for feature extraction in fingerprint images," in *Int. Conference on Bioinformatics and its Applications*, 2004, p. 344.

[34] A. Jain, S. Prabhakar, L. Hong, and S. Pankanti, "Filterbank-based fingerprint matching," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 846–859, May 2000.

[35] J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21–30, Jan. 2004.

[36] B. Krishnapuram, L. Carin, M. Figueiredo, and A. Hartemink, "Sparse multinomial logistic regression: fast algorithms and generalization bounds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 957 –968, June 2005.

[37] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.

[38] K. Nandakumar, Y. Chen, S. Dass, and A. Jain, "Likelihood ratio-based biometric score fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 342–347, Feb. 2008.

[39] X. F. M. Yang, L. Zhang and D. Zhang, "Fisher Discrimination Dictionary learning for sparse representation," *International Conference on Computer Vision*, 2011.